12

Reprinted from *IEEE Transactions on Communication Technology*, Vol. COM-13, No. 2, June 1965

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Report No. 32-787

Optimum Coherent Linear Demodulation

William C. Lindsey

This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

October 1, 1965

Optimum Coherent Linear Demodulation

WILLIAM C. LINDSEY, MEMBER, IEEE

Abstract—**Presented is the performance analysis for four types of linear-modulated communication systems where the message to be transmitted comes from one of two classes of stochastic processes, and the additive channel noise is white and Gaussian. The two classes of stochastic processes which are used to modulate the transmitter are taken to be the "maximally flat" and "asymptotically Gaussian" processes. Demodulation is accomplished at the receiver by coherent frequency translation using a noisy replica of the carrier and filtering the result with one of two types of Wiener filters. These are commonly referred to as the zero-lag (realizable) and infinite-lag (nonrealizable) Wiener filters.**

The four types of modulation considered are: linear-modulation, double-sideband (DSB); linear-modulation, double-sideband, suppressed carrier (DSB/SC); linear-modulation, single-sideband (SSB); and linear-modulation, single-sideband, suppressed carrier (SSB/SC).

Introduction

A PROBLEM of current interest in the area of space communications is that of utilizing one of the neighboring planets as a parasitic antenna for reflecting an analog signal between two widely separated (or the same) points on the earth and then detecting the transmitted signal. The question that immediately comes to mind to the system design engineer is that of selecting the modulation-demodulation technique which allows for the most unambiguous detection procedure at the receiver. In this paper we consider the following four types of linear modulation-demodulation techniques and compare each technique based on two classes of modulating spectra.

Type I: linear-modulated, double-sideband (DSB).

Type II: linear-modulated, double-sideband, suppressed carrier (DSB/SC).

Type III: linear-modulated, single-sideband (SSB).

Type IV: linear-modulated, single-sideband, suppressed carrier (SSB/SC).

Manuscript received July 21, 1964; revised March 25, 1965. This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract NAS 7-100, sponsored by NASA.

The author is with the Communications Systems Research Section of the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif.

The results obtained, however, may be applied to a wide class of problems which requires the transmission of analog information to a distant point, e.g., the transmission of analog data from a vehicle in orbit about the moon.

The communication links under consideration are depicted in Figs. 1 and 2. At the transmitter (*XMTR*) the kth random process of the ith message class $\{m_{ki}(t)\}$, ($i = 1, 2; k = 1, 2, \ldots, \infty$) is used to modulate the transmitter. The output waveform, say $\xi_{ki}(t)$, is transmitted into the channel where additive white Gaussian noise of single-sided spectral density N_0 W/(c/s) corrupts the transmitted waveform resulting in the received waveform $\psi_{ki}(t) = \xi_{ki}(t) + \nu(t)$. The detection procedure is carried out as follows: The observed data $\psi_{ki}(t)$ are multiplied by a noisy copy, say $r(t)$, of the transmitted carrier, and the resulting waveform $\eta_{ki}(t)$ is filtered (after an appropriate transformation in the SSB systems) by one of two types of Wiener filters, i.e., the appropriate linear filter that minimizes the mean-square error, [1]. A filter of type I works as follows. The input function $\eta_{ki}(t)$ [or $x_{ki}(t)$] is recorded for a certain interval of time (theoretically for $-\infty < t < \infty$)

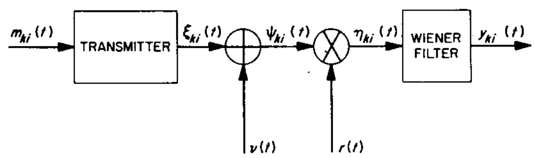

Fig. 1. Communication link—DSB.

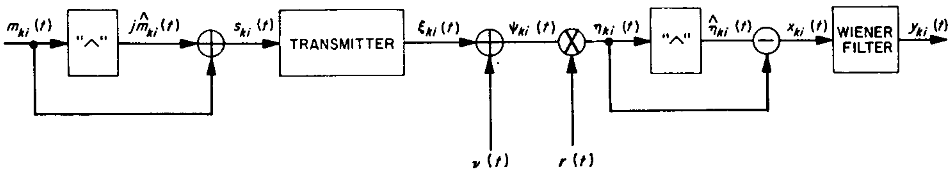

Fig. 2. Communication link—SSB.

and is then processed. For type II filtering we assume that the filter is physically realizable and may be constructed by a circuit containing resistances, inductances, and capacitances. In certain applications, e.g., reflection of the analog signal from a neighboring planet, where delay in the demodulation procedure is of no importance, type I filtering may be practical.

The advantages of type II filters are the simplicity with which they may be implemented and the rapidity with which the output data are delivered. The advantage of type I filters is the more complete use they make of the input signal; consequently, the additive noise may be suppressed more effectively. A comparison of both types of filtering action will be given (for two classes of message spectra) on the basis of a "signal-to-noise ratio" related to the Wiener error vs. a "signal-to-noise ratio" determined by initial design parameters.

The Signaling Processes

At the transmitter we presume we have available two classes of stationary time series with spectral densities denoted by $S_1(\omega; k)$and $S_2(\omega; k)$, $(k = 1, 2 \ldots, \infty)$. Class one is taken to be of the "maximally flat" form, i.e.,

$$S_1(\omega; k) = \frac{K_1(k)}{1 + (\omega/a)^{2k}}; k = 1, 2, \ldots, \infty \tag{1}$$

where $K_1(k)$ is a constant that is chosen so that the time series, which it represents, has unit variance, i.e.,

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} S_1(\omega; k)\, d\omega = 1.$$

For the class of "maximally flat" spectra $K_1(k)$ is given by

$$K_1(k) = (\pi/a) \operatorname{sinc} (\pi/2k) \tag{2}$$

and we have adopted the notation that $\operatorname{sinc} x = \sin x/x$. This process is both physically reasonable, mathematically convenient, and the integer k is a measure of the rate of spectrum cutoff, e.g., $k = 1$ corresponds to a dropoff of 6 dB per octave, $k = 2$ corresponds to 12 dB per octave, etc. Further, $a/2\pi$ may be considered to be the half-power frequency of the time series $m_{ki}(t)$. If $k = 1$, $S_1(\omega; 1)$ is the spectral density occurring at the output of an RC circuit whose input is white Gaussian noise. For $k = \infty$, we have

$$S_1(\omega; \infty) = \begin{cases} \pi/a; & \|\omega\| < a \\ 0; & \|\omega\| > a \end{cases}$$

which is the impulse power response of an ideal low-pass filter of bandwidth $a/2\pi$ c/s.

Class two processes are taken to be the stationary "asymptotically Gaussian" processes with a spectral density given by

$$S_2(\omega; k) = \frac{K_2(k)}{[1 + (\omega/a\sqrt{k})^2]^k}; k = 1, 2, \ldots, \infty \tag{3}$$

and $K_2(k)$ is adjusted such that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} S_2(\omega; k)\, d\omega = 1.$$

Direct substitution of (3) into this expression yields the value

$$K_2(k) = \frac{4\pi}{a\sqrt{k}\mathrm{B}(\frac{1}{2}, k - \frac{1}{2})} \tag{4}$$

where $\mathrm{B}(u, \nu)$ is the well-known beta function. If $k = 1$, $S_1(\omega; 1) = S_2(\omega; 1)$, while as k approaches infinity in (3) we have

$$S_2(\omega; k) = K_2(k)\left[1 + \left(\frac{\omega}{a\sqrt{k}}\right)^2\right]^{-k}$$
$$= K_2(k) \exp\left[-k\left(\frac{(\omega/a)^2}{k} \frac{(\omega/a)^4}{2k^2} + \frac{\omega/a)^6}{3k^3} - \ldots\right)\right]$$

or

$$\lim_{k \to \infty} S_2(\omega; k) = \frac{2\sqrt{\pi}}{a} \exp\left[-\left(\frac{\omega}{a}\right)^2\right] \tag{5}$$

which is the Gaussian spectrum. This unit variance process is rather interesting from the physical standpoint in that it may be generated by passing white Gaussian noise through k isolated cascaded RC networks. Note that the two random processes possess radically different frequency components as k becomes large. For $k = \infty$, the parameter $a/2\pi$ may be considered to be that frequency at which the spectrum has decayed to e^{-1} times the value at $\omega = 0$. These two classes of random processes are sufficiently general in that they include a broad class of signaling spectra encountered in communication engineering.

The Transmitted Signals

A representation of these signals which is most convenient for our purposes is to represent the transmitted waveforms as the product of a real low-pass waveform, which depends on the modulating signal $m_{ki}(t)$ and a complex cisoidal carrier. For type I modulation, i.e., DSB, the transmitted signal may be written as

$$\xi_{ki}(t) = \sqrt{2P}[1 + m_a m_{ki}(t)] \exp (j\omega_c t); \; \begin{matrix} i = 1, 2 \\ k = 1, 2, \ldots, \infty \end{matrix} \tag{6}$$

where ω_c is a suitably defined carrier frequency and 100 m_a is a measure of the per cent of modulation. It is the real part of (6) which corresponds to the actual transmitted signal. The autocorrelation function $R_1(\tau)$ of (6) is

$$R_1(\tau) = 2P[1 + m_a{}^2 R_{m_{ki}}(\tau)] \exp (j\omega_c\tau) \tag{7}$$

where $R_{m_{ki}}(\tau)$ is the autocorrelation function of the kth member of the ith stochastic process. It may be shown [2] that one-half times the real part of (7) represents the autocorrelation function of the actual transmitted signal. Since $R_{m_{ki}}(0) = 1$ for all i and k, the actual total average power transmitted is given by

$$P_t = \tfrac{1}{2}R_1(0) = P[1 + m_a{}^2]. \tag{8}$$

The DSB/SC signal may be written mathematically as

$$\xi_{ki}(t) = \sqrt{2P} m_{ki}(t) \exp (j\omega_c t); \; \begin{matrix} i = 1, 2 \\ k = 1, 2 \ldots, \infty \end{matrix} \tag{9}$$

where we have assumed that the kth member of the ith stochastic signaling class is being transmitted. The autocorrelation function of (9) is given by

$$R_2(\tau) = 2PR_{m_{ki}}(\tau) \exp (j\omega_c\tau) \tag{10}$$

and, since $R_{m_{ki}}(0)=1$,

$$P_t = \tfrac{1}{2}R_2(0) = P \tag{11}$$

is the total average power transmitted.

The SSB (type III) signal is a bit more difficult to generate. In this case we transmit

$$\xi_{ki}(t) = \sqrt{2P}[1 + m_a s_{ki}(t)] \exp(j\omega_c t); \quad i = 1, 2; \; k = 1, 2, \ldots, \infty \tag{12}$$

where the process $s_{ki}(t)$ is generated in the following manner.

Assume that the kth member of the ith process is being transmitted. The time series $m_{ki}(t)$ is passed through a *Hilbert transforming* ("$\hat{\ }$") filter whose output has been phase-shifted by 90° and represented by the waveform $\hat{m}_{ki}(t)$. This process is added to the original process to produce the signal

$$s_{ki}(t) = m_{ki}(t) + j\hat{m}_{ki}(t) \tag{13}$$

where ("$\hat{\ }$") signifies the Hilbert transforming operation (see Fig. 2). The spectral properties of $s_{ki}(t)$ may be shown [2] to be related to those of $m_{ki}(t)$ through

$$S_{s_{ki}}(\omega) = \begin{cases} 4S_i(\omega; k); & \omega > 0 \\ 2S_i(\omega; k); & \omega = 0. \\ 0 & ; \omega < 0 \end{cases} \tag{14}$$

It may also be shown that [2]

$$R_m(\tau) = R_{\hat{m}}(\tau) \tag{15}$$

and

$$E[m(\tau)\hat{m}(s)] = E[m(\tau)]E[\hat{m}(s)] \tag{16}$$

which says the process $m(\tau)$ and its Hilbert transform are uncorrelated. If we further assume that both signaling processes have zero mean, it follows that the autocorrelation function of (12) is

$$R_3(\tau) = 2P[1 + m_a^2 R_{s_{ki}}(\tau)] \exp(j\omega_c\tau) \tag{17}$$

where use has been made of (12) and (16). Thus, for the SSB signal the total average transmitted power is

$$P_t = \tfrac{1}{2}R_3(0) = [1 + 2m_a^2]P. \tag{18}$$

For type IV modulation, i.e., SSB/SC, we transmit

$$\xi_{ki}(t) = \sqrt{2P}s_{ki}(t) \exp(j\omega_c t) \tag{19}$$

where $s_{ki}(t)$ is given by (13). The autocorrelation of (19) is easily shown to be

$$R_4(\tau) = 2PR_{s_{ki}}(\tau) \exp(j\omega_c\tau) \tag{20}$$

and the average total transmitted power is given by

$$P_t = \tfrac{1}{2}R_4(0) = 2P. \tag{21}$$

The real part of (6), (9), (12), and (19) represents the physical signal emitted by the transmitter when the average power in the carrier is the same for all four types of modulation techniques. If, on the other hand, we have available for transmission only P_t watts, regardless of the type of modulation technique, we may write from (8), (11), (18), and (21) the set of "normalized" transmitted signals $\{\xi_{ki}(t)\}$ using complex carriers as

$$\begin{aligned}
\xi_{ki}(t) &= \sqrt{\frac{2P_t}{1 + m_a^2}}\,[1 + m_a m_{ki}(t)] \exp(j\omega_c t); \text{ DSB} \\
\xi_{ki}(t) &= \sqrt{2P_t}m_{ki}(t) \exp(j\omega_c t); \text{ DSB/SC} \\
\xi_{ki}(t) &= \sqrt{\frac{2P_t}{1 + 2m_a^2}}\,[1 + m_a s_{ki}(t)] \exp(j\omega_c t); \text{ SSB} \\
\xi_{ki}(t) &= \sqrt{P_t}s_{ki}(t) \exp(j\omega_c t); \text{ SSB/SC}
\end{aligned} \tag{22}$$

where we have assumed we are transmitting the kth member of the ith message class, $i=1, 2$; $k=1, 2, \ldots, \infty$. This is sufficient to characterize the transmitted waveforms. Practical methods for impressing the signals m_{ki} onto the carrier, i.e., generation of the real parts of (22), are given in [3] and [4] while methods of signal reception are given in Coatas [5] and Norgaard [6].

Characterization of the Additive Noise

We presume that the complex additive noise $\nu(t)$ is given by

$$\nu(t) = n_1(t) \exp(j\omega_c t) \tag{23}$$

where

$$n_1(t) = n(t) + j\hat{n}(t)$$

and $n(t)$ and its Hilbert transform $\hat{n}(t)$ are white Gaussian noise processes possessing single-sided spectral densities of N_0 W/cps. The physical additive noise process is the "real part" of the complex Gaussian process $\nu(t)$, i.e.,

$$n_0(t) = n(t) \cos \omega_c t - \hat{n}(t) \sin \omega_c t. \tag{24}$$

The autocorrelation function of $n_0(t)$ is easily shown to be

$$R_{n_0}(\tau) = (N_0/2)\delta(\tau).$$

In carrying out the frequency-translation operation at the receiver, one must use the real part of the received signals $\psi_{ki}(t)$. If we multiply the noise process $n_0(t)$ by the noisy stored carrier reference $r(t)$, we obtain

$$n_0(t)r(t) = (1/\sqrt{2})[n(t) \cos \Phi + \hat{n}(t) \sin \Phi] + \text{double frequency terms}$$

where we have assumed that the stored reference is given by

$$r(t) = \sqrt{2} \cos(\omega_c t + \Phi) \tag{25}$$

and Φ is a random variable. For example, Φ may well represent the phase error of a phased-locked loop which is tracking the sinusoid $\sin \omega_c t$ in the presence of additive white Gaussian noise. Several probability distributions have been derived in [7] and [8], which govern the statistics of this phase error. Neglecting the double-frequency terms (since the Wiener filter will not respond to them), we may represent the noise at the multiplier output by

$$n'(t) = (1/\sqrt{2})[n(t) \cos \Phi + \hat{n}(t) \sin \Phi]. \tag{26}$$

It remains to determine the statistics of $n'(t)$. The autocorrelation function of the noise process $n'(t)$ may be shown to be

$$\begin{aligned} R_{n'}(t-s) &= E[n'(t)n'(s)] \\ &= (N_0/2)\delta(t-s) \end{aligned} \tag{27}$$

by using (26) and the facts that n and $\hat{n}$ are uncorrelated and have zero mean. Equation (27) says that the noise present at the multiplier output is white and Gaussian and has a single-sided spectral density of N_0 W/(c/s).

For reasons that will become obvious later, we compute the multiplier outputs $\eta_{ki}(t)$ for all four types of modulation. To accomplish frequency translation in the physical sense we must use the physical waveforms received, i.e., $\text{Re}\{\psi_{ki}(t)\}$, where Re denotes "real part." For the DSB systems we have, using (22), (24), (26), and a little labor,[1]

$$\begin{aligned} \eta_{ki}(t) &= m_a\sqrt{P_t/1 + m_a^2}m_{ki}(t) + n'(t); \text{DSB} \\ \eta_{ki}(t) &= \sqrt{P_t}m_{ki}(t) + n'(t); \text{DSB/SC} \end{aligned} \tag{28}$$

and we have neglected the double-frequency and dc terms. In the SSB cases we have for the multiplier outputs, neglecting the double-frequency and dc terms,

$$\begin{aligned} \eta_{ki}(t) &= m_a\sqrt{\frac{P_t}{1+2m_a^2}}[m_{ki}(t) + \hat{m}_{ki}(t)] + n''(t)/\sqrt{2}; \text{SSB} \\ \eta_{ki}(t) &= \sqrt{\frac{P_t}{2}}[m_{ki}(t) + \hat{m}_{ki}(t)] + \frac{n''(t)}{\sqrt{2}}; \text{SSB/SC} \end{aligned} \tag{29}$$

where $n''(t) = n(t) + \hat{n}(t)$. The input to the Wiener filter for the SSB systems is $x_{ki}(t) = \eta_{ki}(t) - \hat{\eta}_{ki}(t)$ or

$$\begin{aligned} x_{ki}(t) &= \sqrt{2}\left[m_a\sqrt{2P_t/1 + 2m_a^2}m_{ki}(t) + n(t)\right]; \text{SSB} \\ x_{ki}(t) &= \sqrt{2}\left[\sqrt{P_t}m_{ki}(t) + n(t)\right]; \text{SSB/SC} \end{aligned} \tag{30}$$

where $n(t)$ is white Gaussian noise of single-sided spectral density N_0 W/(c/s). Equations (28) and (30) represent, respectively, the inputs (signal plus noise) to the Wiener filters of Figs. 1 and 2. Note that, for the DSB/SC and SSB/SC systems the signals to be filtered are essentially the same since the square root of two in (30) may be neglected because it effects both the signal and noise. The spectral densities of the input process for the four types of modulation may be written from (1), (3), (28), and (30). Neglecting the square root of two in (30) they are

$$\begin{aligned} S_{1i}(\omega; k) &= \delta_1 S_i(\omega; k); \text{DSB} \\ S_{2i}(\omega; k) &= \delta_2 S_i(\omega; k); \text{DSB/SC} \\ S_{3i}(\omega; k) &= \delta_3 S_i(\omega; k); \text{SSB} \\ S_{4i}(\omega; k) &= \delta_4 S_i(\omega; k); \text{SSB/SC} \end{aligned} \tag{31}$$

for $i = 1, 2$, and all k. The δ_j factors are defined as

$$\begin{aligned} &\delta_1 = m_a^2(1 + m_a^2)^{-1}P_t = g_1P_t;\ \delta_2 = g_2P_t = P_t \\ &\delta_3 = 2m_a^2(1 + 2m_a^2)^{-1}P_t = g_3P_t;\ \delta_4 = g_4P_t = P_t \end{aligned} \tag{32}$$

while the average input signal power is

$$P_j = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{ji}(\omega; k)d\omega; j = 1, 2, 3, 4$$

which becomes, by using (31) and (32),

$$\begin{aligned} &P_1 = m_a^2(1 + m_a^2)^{-1}P_t;\ P_2 = P_t \\ &P_3 = 2m_a^2(1 + 2m_a^2)^{-1}P_t;\ P_4 = P_t. \end{aligned} \tag{33}$$

The Wiener Error for the Two Classes of Signaling Spectrums

The instantaneous value of the Wiener error may be written assuming the kth member of the ith stochastic class is being transmitted using the jth modulation technique as (see Figs. 1 and 2)

$$\epsilon_{ki}^{j}(t) = y_{ki}(t) - \sqrt{\delta_j}m_{ki}(t); j = 1, 2, 3, 4$$

where the δ_j's are defined in (32). Since the modulating signal and the noise vary randomly with time, it is natural to characterize the "output noise" by its mean square intensity

$$\overline{(\epsilon_{ki}^{j})^2} = \overline{\left[y_{ki}(t) - \sqrt{\delta_j}m_{ki}(t)\right]^2}.$$

The Wiener filter (types I and II), which minimizes the mean-square error for all members of the two classes of stochastic processes, is the filter that we use at the receiver for smoothing the observed data $\eta_{ki}(t)$ and $x_{ki}(t)$. The filter functions (impulse responses) are determined from the spectral densities of the signal and noise; however, we are not interested here in the frequency responses of the individual filters.

Instead we shall be concerned primarily with determining the filtering action, i.e., computation of the Wiener error $\overline{(\epsilon_{ki}^{j})^2} = \sigma_{ji}^2(k)$.

For type II filters (nonrealizable) it may be shown that the mean-square error occurring when one transmits the kth member of the ith signal class using the jth modulation technique is given by [1],

$$\sigma_{ji}^2(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_{ji}(\omega; k)}{1 + [2S_{ji}(\omega; k)]/N_0}d\omega \tag{34}$$

and we have assumed the input noise is white and $S_{ji}(\omega; k)$ are the spectral densities given by (31).

On the other hand, for type I filters (realizable), the mean-square error encountered (when the kth member of

[1] At this point we have assumed perfect coherence at the receiver. The noisy phase reference case will be discussed later. In the SSB systems we have assumed that $r(t) = \text{Re}[\sqrt{2}(1 + j1)\exp(j\omega_c t)]$. Since $\hat{m}$ and m have equivalent spectral densities, this serves to illustrate how either m or $\hat{m}$ may be recovered at the receiver and filtered with the same Wiener filter. If $\hat{m}$ is desired, one must form $\eta + \hat{\eta}$ instead of the difference.

the ith signal class is transmitted and the jth modulation technique employed at the transmitter) may be obtained from (34) by multiplying the spectral density in the denominator by x and integrating with respect to x over the unit interval (0, 1). This procedure yields, for white noise with a single-sided spectral density of N_0 W/(c/s),

$$\sigma_{ji}^2(k) = \frac{N_0}{4\pi}\int_{-\infty}^{\infty} \ln\left[1 + \frac{2S_{ji}(\omega; k)}{N_0}\right] d\omega \quad (35)$$

which is the Wiener error obtained by Yovits and Jackson [9]. Equations (34) and (35) are remarkable in that the Wiener error may be evaluated without having to compute the individual filter functions.

Performance of Types I and II Filters Using "Maximally Flat" Spectra

The Wiener error for all four types of modulation and all members of both signaling classes may be computed from (31) and (34), i.e.,

$$\sigma_{ji}^2(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\delta_j S_i(\omega; k)}{1 + (2/N_0)\delta_j S_i(\omega; k)}\, d\omega; \quad i = 1, 2 \quad j = 1, 2, 3, 4 \quad k = 1, 2, \ldots, \infty \quad (36)$$

where the constants δ_j and g_j are defined in (32). Letting $i=1$ in (36) and substituting (1) into (36), it may be shown (see Appendix I) that

$$\sigma_{j1}^2(k) = \delta_j[1 + 2\delta_j K_1(k)/N_0]^{(1/2k)-1} \quad (37)$$

where $K_1(k)$ is given by (2). Defining the SNR ρ as the ratio of the mean-squared value of the signal power P_j [see (33)] to the Wiener error, we have[2]

$$\rho_{j1}(k; I) = \left[1 + 2\pi g_j R \operatorname{sinc}\left(\frac{\pi}{2k}\right)\right]^{1-(1/2k)} \quad (38)$$

where $R = P_t/aN_0$ and the g_j's are given by (32). For the "maximally flat" case we have for $k = \infty$

$$\rho_{j1}(\infty; \mathrm{I}) = 1 + 2\pi g_j R \quad (39)$$

while for large values of the parameter R, (39) becomes

$$\rho_{ji}(k; \mathrm{I}) \sim [2\pi g_j R \operatorname{sinc}(\pi/2k)]^{1-(1/2k)}. \quad (40)$$

The Wiener error for type II filters operating on signal class one is shown in Appendix II to be

$$\sigma_{j1}^2(k) = \frac{kN_0}{K_1(k)}\{[1 + 2\delta_j K_1(k)/N_0]^{1/2k} - 1\}.$$

The SNR ρ becomes

$$\rho_{j1}(k; \mathrm{II}) = \frac{\pi g_j R \operatorname{sinc}(\pi/2k)}{k\{[1 + 2\pi g_j R \operatorname{sinc}(\pi/2k)]^{1/2k} - 1\}} \quad (41)$$

which for large R is asymptotic to

$$\rho_{j1}(k; \mathrm{II}) \sim \frac{\pi g_j R \operatorname{sinc}(\pi/2k)}{k\{[2\pi g_j R \operatorname{sinc}(\pi/2k)]^{1/2k} - 1\}}. \quad (42)$$

Of special interest is the case where $k = \infty$. It may be shown that

$$\rho_{j1}(\infty; \mathrm{II}) = 2\pi g_j R/\ln[1 + 2\pi g_j R] \quad (43)$$

which for large R is asymptotic to

$$\rho_{j1}(\infty; \mathrm{II}) \sim 2\pi g_j R/\ln(2\pi g_j R). \quad (44)$$

Comparison of (40) with (42) shows that, for large R and small k, type I filters have an SNR of approximately 2 k times the SNR of type II filters. As k approaches infinity, (39) and (44) show that the performance of type II filtering becomes inferior to type I filtering by a factor of ln $(2\pi g_j R)$. For $k=1$ and large R, type I filters outperform type II filters by a factor approximately 3 dB.

Performance of Types I and II Filters Using "Asymptotically Gaussian" Spectra

Equation (3) may be substituted into (31) yielding the spectra for the four types of modulation. This result, when used in (34) and (35), gives the required Wiener error. Because of the lengthy details and integration procedure required for general k, we evaluate the SNR ρ for the special case $k=\infty$.

For type II filtering the Wiener error for the jth modulation technique is given by

$$\sigma_{j2}^2(\infty) = \frac{aN_0}{4\pi}\int_{-\infty}^{\infty} \ln\left[1 + 4\sqrt{\pi} g_j R \exp(-x^2)\right] dx$$

and the SNR ρ becomes

$$\rho_{j2}(\infty; \pi) = \frac{2\pi g_j R}{\int_0^{\infty} \ln\left[1 + 4\sqrt{\pi} g_j R \exp(-x^2) dx\right]} \quad (45)$$

For type I filtering the Wiener error for the jth type of modulation is easily shown to be

$$\sigma_{j2}^2(\infty) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{2\sqrt{\pi} g_j P_t \exp(-x^2)}{1 + 4\sqrt{\pi} g_j R \exp(-x^2)}\, dx$$

and the SNR becomes

$$\rho_{j2}(\infty; \mathrm{I}) = \frac{\sqrt{\pi}/2}{\int_0^{\infty} \frac{\exp(-x^2) dx}{1 + 4 g_j \sqrt{\pi} R \exp(-x^2)}} \quad (46)$$

Equations (45) and (46) may be integrated by expanding the integrand into an infinite series and integrating term by term. Difficulty arises, however, when $4g_j\sqrt{\pi}R > 1$. A more tractable procedure to use is to integrate (45) and (46) numerically on a general purpose computer.

[2] The notation $\rho_{ji}(k; I)$ signifies the SNR when the kth member of the ith stochastic process is being transmitted using the jth modulation technique, and type I filters are used at the receiver.

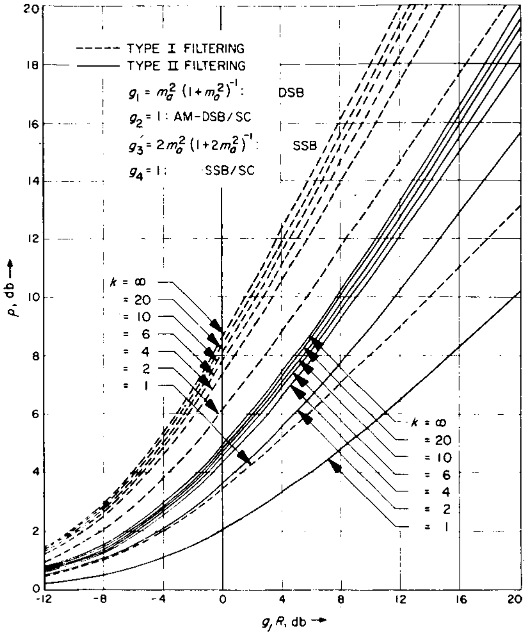

Fig. 3. System performance characteristics.

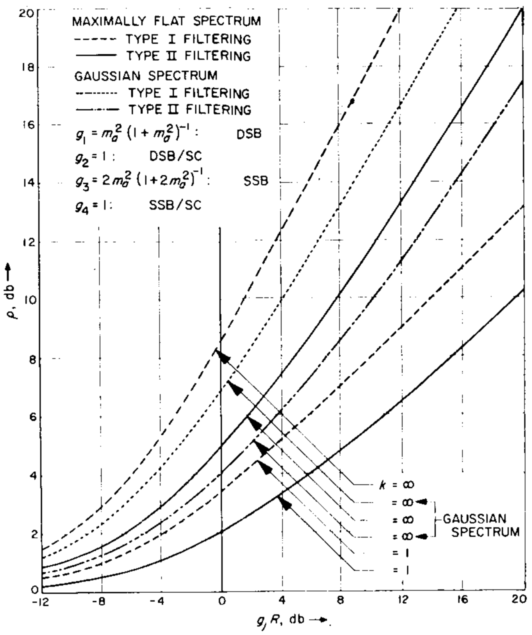

Fig. 4. System performance characteristics.

Calculated Performance and Comparison

Plotted in Figs. 3 and 4 is the SNR ρ vs. the basic parameter g_jR, where $R = P_t/aN_0$. In particular, Figs. 3 and 4 have been plotted for $g_2 = g_4 = 1$, i.e., DSB/SC and SSB/SC systems. Performance for the other two types of modulation may be obtained from these figures by rescaling the abscissa by g_j, e.g., if the performance of the DSB system is required, $j = 1$ and $g_1 = m_a^2(1 + m_a^2)^{-1}$ [see (32)].

The curves show that, regardless of the type of amplitude modulation employed at the transmitter, the larger k (for either class of stochastic signals), the better is the SNR ρ. This is easily explained on a physical basis. For large k, energy in the signaling spectra is suppressed in the high-frequency regions and accentuated in the low-frequency regions. Hence, the Wiener filter, for a white noise input, accepts a smaller amount of the input noise, and the SNR ρ is larger. Note that, for large k and R, type I filters (nonrealizable) yield a value of ρ highly superior to type II filters (realizable). In physical situations where delay in the demodulation procedure is tolerable, it is quite evident that type I filtering should be employed with either the SSB/SC or the DSB/SC system. If bandwidth is a premium, then the SSB/SC system should be selected over the DSB/SC system. All systems have the disadvantage of requiring a local copy of the carrier at the receiver. For SSB and DSB, a carrier component is available in the observed data. Such is not the case for SSB/SC and DSB/SC; consequently, other means must be employed for obtaining this information at the receiver. This is obviously a disadvantage of either of these systems over the DSB and SSB systems. In terms of transmission bandwidth, DSB and DSB/SC require equal amounts while SSB and SSB/SC require only half as much as the DSB or DSB/SC system.

If we view the parameter $g_jR = g_jP_t/aN_0$ as a measure of the effectiveness of the jth modulation technique, we find that DSB/SC and SSB/SC perform equally well. On this basis, the DSB/SC and SSB/SC systems are $10 \log_{10} [m_a^{-2}(1 + m_a^2)]$ better than DSB and $10 \log_{10} [(1 + 2m_a^2)(2m_a^2)^{-1}]$dB better than SSB. In terms of ρ, for a given k and R, no general conclusions may be reached; the curves in Figs. 3 and 4 must be consulted.

Performance Using a Noisy Phase Reference

One major difficulty with implementing any of the amplitude-demodulation systems studied here is that of providing the receiver with a copy of the transmitted carrier, i.e., synchronization of the transmitter and receiver local oscillators. One practical means of achieving carrier synchronization in the past has been to employ a phase-locked loop at the receiver. Even if one is willing to build a phase-locked loop at the receiver, there remains the question of what component in the received signal should one try to achieve carrier lock; e.g., in DSB/SC or SSD/SC, the received spectrum does not contain a frequency component oscillating at the carrier frequency. The best one can do (probably) is to transmit a pilot carrier for use in connection with the phase-locked loop. Other methods are outlined in [3], [5], and [6]. This, however, requires additional energy. On the other hand, for DSB and SSB there exists a carrier component in the received signal spectrum. In fact, the power in this component is a function of the modulation index m_a [see (22)]. With a knowledge of this it is not at all clear how one could most effectively mechanize a phase-locked loop for synchronization purposes. Consequently, in what follows we shall assume that a phase-locked loop is used to derive the reference signal

$$r(t) = \sqrt{2}\cos(\omega_c t + \Phi) \tag{47}$$

where Φ is the phase error, and the voltage control oscillator (VCO) in the phase-locked loop (PLL) is oscillating at the carrier frequency. Viterbi [7] and Tikhonov [8] have shown that the probability distribution $p(\Phi)$ for the phase error Φ is given by

$$p(\Phi) = \frac{1}{2\pi I_0(\alpha)} \exp(\alpha \cos \Phi) \tag{48}$$

where $I_0(\alpha)$ is the imaginary Bessel function evaluated at the SNR existing in the loop. Taking into consideration a phase error Φ rad/s, the factor $\cos \Phi$ multiplies the signal components in (28) and (30). As already shown, the noise statistics remain unchanged. Hence, the spectral densities of (31) are multiplied by $\cos^2 \Phi$ as well as the g_j's of (32) and the Wiener error; computed for the ideal reference signal, they become that value of the Wiener error, conditional on the fact that the phase error is Φ rad/s, i.e., $\sigma_{ji}(k) = \sigma_{ji}(k|\Phi)$. The total mean-square error that results when all members of the phase-error ensemble are taken into consideration becomes

$$\sigma_{ji}^2(k) = \int_{-\pi}^{\pi} p(\Phi)\sigma_{ji}^2(k|\Phi)d\Phi. \tag{49}$$

If one attempts to solve this equation using (48) for general k, j, and i, a formidable integral is immediately encountered. Special cases, e.g., $k = \infty$, may be worked out exactly. For general j, k, and i, numerical integration techniques could be applied to obtain values for ρ_{ji} (k; I or II), but it appears, at this point, to be hardly worth the effort.

An alternate procedure that gives some idea as to the effect of a noisy phase reference is to average over the phase error before filtering, i.e., define the input signal component of $\eta_{ki}(t)$ [or $x_{ki}(t)$] by the following relationship.

$$\eta_{ki}(t) = \int_{-\pi}^{\pi} p(\Phi)\eta_{ki}(t/\Phi)d\Phi. \tag{50}$$

Carrying out this integral using (28) and (48) yields for the jth type of modulation

$$\eta_{ki}(t) = \frac{1}{2\pi I_0(\alpha)}\left[\int_{-\pi}^{\pi} \sqrt{g_j}\, P_t m_{ki}(t) \cos \Phi \exp(\alpha \cos \Phi) d\Phi + \int_{-\pi}^{\pi} n'(t) \exp(\alpha \cos \Phi) d\Phi\right]. \tag{51}$$

Performing the integration gives

$$\eta_{ki}(t) = \sqrt{g_j}\frac{I_1(\alpha)}{I_0(\alpha)} P_t m_{ki}(t) + n'(t) = \sqrt{g_j'}\, P_t m_{ki}(t) + n'(t) \tag{52}$$

where

$$\sqrt{g_j'} = \sqrt{g_j}\,[I_1(\alpha)/I_0(\alpha)]$$

and $I_1(\alpha)$ is the first-order Bessel function of imaginary argument. Thus, for all types of modulation the curves of Figs. 3 and 4 still apply; however, the abscissa is now $g_j'R$ instead of g_jR. If $\alpha = \infty$, corresponding to perfect coherence, $g_j' = g_j$. If $\alpha = 0$, corresponding to a carrier whose phase variable is uniformly distributed over an interval of length 2π, $g_j' = 0$ for all j. For $0 < \alpha < \infty$ we find that $g_j' < g_j$, e.g., if $\alpha = 10$, corresponding to an SNR in the loop of 10 dB, we have $g_j' = 0.90g_j$. If $\alpha = 1$, corresponding to an SNR in the loop of zero dB, we find $g_j' = 0.20g_j$. Thus we see that a good (non-noisy) replica of the carrier is required at the receiver in order that the demodulation procedure be performed efficiently.

Conclusions

In this paper we have analyzed four types of amplitude modulation-demodulation systems. The information-bearing signal used to modulate the transmitter is generated from one of two classes of stochastic processes; the "maximally flat" and the "asympotically Gaussian" processes. We have shown that system performance depends on the type of Wiener filter (realizable or nonrealizable) used to smooth the noisy data and the modulating spectrum. In particular, for $k = 1$ and large R, the nonrealizable filter performs approximately 3 dB better than the realizable filter. For large k and R the nonrealizable filter performs approximately $10 \log_{10} [\ln 2\pi g_j R]$ dB better than the realizable filter.

It is shown that system performance is highly dependent on the parameter k of the modulation spectrum. In fact, it is advisable to shape the modulating spectrum before transmission by means of a Butterworth filter or a series of isolated-cascaded RC networks. Shaping of the modulation spectrum by a Butterworth filter proves to be more effective than that of using a series of isolated-cascaded RC networks.

Finally, we considered the situation where the receiver utilizes in the demodulation procedure a noisy replica of the transmitted carrier. If the carrier replica is derived at the receiver by means of a phase-locked loop we find that the Wiener error is least when the carrier replica is relatively noise free. For example, a SNR of 10 dB in the tracking loop reduces the effective input SNR at the demodulator input by 0.4 dB whereas an SNR of zero dB in the tracking loop reduces the effective input SNR by 7 dB. Finally, we point out that these results are compared in [10] with similar results obtained for frequency demodulation using phase-locked frequency discriminators.

Appendix I

We wish to derive (38). From (1) and (36) we may write

$$\sigma_{ji}^2(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\dfrac{\delta_j K_1(k) d\omega}{1 + (\omega/a)^2}}{1 + \dfrac{\delta_j K_1(k)}{1 + (\omega/a)^{2k}}}.$$

Rearranging and making the change of variable $ax = \omega$

yields

$$\sigma_{j1}^2(k) = \delta_j \frac{aK_1(k)}{\pi} [1 + 2\delta_j K_1(k)/N_0]^{(1/2k)-1} \int_0^\infty \frac{dx}{1+x^{2k}}.$$

Performing the integration and simplifying gives the desired result.

Appendix II

To derive (41) we make use of Leibnitz's rule for integrals which depend on a parameter, say x. Leibnitz's rule says that, if the function $f(x, \omega)$ is continuous and has continuous derivative $\partial f/\partial x$ in a domain of the x-ω plane, then

$$\frac{d}{dx}\int_a^b f(x, \omega)d\omega = \int_a^b \frac{\partial f}{\partial \omega}(x, \omega)d\omega. \tag{53}$$

In other words, differentiation and integration may be interchanged. Using Leibnitz's rule we present two methods for evaluating integrals involving the logarithmic.

Consider the functions $f(x, \omega) = [1 + h(\omega)]^x$,

$$g(x) = \int_0^\infty [1 + h(\omega)]^x d\omega \tag{54}$$

and the derivative $g'(x) = dg/dx$, i.e.,

$$\frac{dg(x)}{dx} = \frac{d}{dx}\int_0^\infty [1 + h(\omega)]^x d\omega. \tag{55}$$

Applying Leibnitz's rule we have

$$g'(x) = \frac{d}{dx}\int_0^\infty [1 + h(\omega)]^x d\omega = \int_0^\infty \frac{\partial}{\partial x}[1 + h(\omega)]^x d\omega$$

$$g'(x) = \int_0^\infty [1 + h(\omega)]^x \ln [1 + h(\omega)]d\omega.$$

Evaluating $g'(x)$ at $x = 0$ gives

$$g'(0) = \int_0^\infty \ln [1 + h(\omega)]d\omega \tag{56}$$

which is the form of the integral given in (35).

The second method that may be used for evaluating (35) is to apply Leibnitz's rule to the following integral, i.e.,

$$g(x) = \frac{d}{dx}\int_0^\infty \ln [1 + xh(\omega)]d\omega$$

$$= \int_0^\infty \frac{\partial}{\partial x}\ln [1 + xh(\omega)]d\omega \tag{57}$$

where $f(x, \omega) = \ln [1 + xh(\omega)]$. Thus, (6) becomes

$$g(x) = \int_0^\infty \frac{h(\omega)d\omega}{1 + xh(\omega)} \tag{58}$$

and integrating both sides with respect to x over the unit interval gives

$$\int_0^1 g(x)dx = \int_0^1 \int_0^\infty \frac{h(\omega)d\omega dx}{1 + xh(\omega)}$$

$$= \int_0^\infty \ln [1 + h(\omega)]d\omega \tag{59}$$

which is the required form. Thus we evaluate (7) and then integrate the result over the unit interval (0, 1) with respect to the parameter x. In either method, $h(\omega)$ is related to the message spectral density through

$$h(\omega) = S_{j1}(\omega; k).$$

Substituting for $h(\omega)$ and $S_{j1}(\omega; k)$ into (7) gives

$$g(x) = \int_0^\infty \frac{\dfrac{\delta_j K_1(k)d\omega}{1 + (\omega/a)^{2k}}}{1 + \dfrac{2x\delta_j K_1(k)}{N_0[1 + (\omega/a)^{2k}]}}. \tag{60}$$

Carrying out the integration yields

$$g(x) = \delta_j(1 + 2x\delta_j K_1(k)/N_0)^{(1/2k)-1}. \tag{61}$$

The error $\sigma_{j1}^2(k)$ is easily related to $g(x)$ through

$$\sigma_{j1}^2(k) = \frac{N_0}{2\pi}\int_0^1 g(x)dx.$$

Substituting for $g(x)$ and integrating gives

$$\sigma_{j1}^2(k) = \frac{kN_0}{K_1(k)}\left\{\left[1 + 2\pi g_j R \operatorname{sinc}\left(\frac{\pi}{2k}\right)\right]^{(1/2k)} - 1\right\} \tag{62}$$

and the SNR ρ is

$$\rho_{j1}(k; \text{II}) = \frac{g_j P_t}{\sigma_{j1}^2(k)} \tag{63}$$

which is the desired result.

References

[1] Wainstein, L. A., and V. D. Zubakov, *Extraction of Signals from Noise.* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

[2] Deutsch, R., *Nonlinear Transformations of Random Processes.* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

[3] Schwartz, M., *Information Transmission Modulation and Noise.* New York: McGraw-Hill, 1959.

[4] Norgaard, D. E., The phase-shift method of single-sideband signal generation, *Proc. IRE*, vol 44, Dec 1956, pp 1718–1734.

[5] Coatas, J. P., Synchronous communications, *Proc. IRE*, vol 44, Dec 1956, pp 1713–1718; also Some notes on space communications, *Proc. IRE*, vol 47, Aug 1959, pp 1304–1385.

[6] Norgaard, D. E., The phase-shift method of single-sideband signal reception, *Proc. IRE*, vol 44, Dec 1956, pp 1735–1744.

[7] Viterbi, A. J., Phase-locked loop dynamics in the presence of noise by Fokker-Planck techniques, *Proc. IEEE*, vol 51, Dec 1963, pp 1737–1753.

[8] Tikhonov, V. I., Phase-lock automatic frequency control operation in the presence of noise, *Avtomatika i Telemekhanika*, vol 21, Mar 1961, pp 301–309.

[9] Yovits, M. C., and J. L. Jackson, Linear filter optimization with game theory considerations, *IRE Nat'l Conv. Rec.*, pt 4, 1955, pp 193–199.

[10] Lindsey, W. C., Optimum and sub-optimum frequency demodulation, TR 32-637, Jet Propulsion Lab., Pasadena, Calif., May 1964.

Reprinted from IEEE TRANSACTIONS ON COMMUNICATION TECHNOLOGY
Vol. COM-13, No. 2, June 1965
pp. 183–190